vehicles

MDPI

Article

Determinants of Test-to-Reality CO_2 Gaps in European PHEVs: The Limited Role of Battery Capacity

Maksymilian Mądziel [1,*], Paulina Kulasa [1] and Tiziana Campisi [2]

[1] Faculty of Mechanical Engineering and Aeronautics, Rzeszow University of Technology, 35-959 Rzeszow, Poland; paulinakulasa@prz.edu.pl
[2] Department of Engineering and Architecture, Kore University of Enna, Cittadella Universitaria, 94100 Enna, Italy; tiziana.campisi@unikore.it
* Correspondence: mmadziel@prz.edu.pl

Abstract

Plug-in hybrid electric vehicles (PHEVs) are expected to reduce fleet CO_2 emissions, but real-world operation often differs markedly from type-approval values. Using European OBFCM data for 457,555 PHEVs (2021–2023) from 14 manufacturers, we quantify the "test-to-reality" CO_2 gap and assess whether traction battery capacity contains an independent signal or mainly reflects vehicle segmentation and in-use behavior. Battery capacity shows only limited standalone explanatory power, while controlling for segment, monitoring year, and manufacturer and incorporating OBFCM-derived usage indicators greatly improves model fit and substantially reduces the apparent battery–gap relationship. We further find strong heterogeneity across vehicle segments, indicating that battery size is not a universal lever of real-world PHEV CO_2 performance. Overall, the results support interpreting battery capacity primarily as a proxy for market positioning and real-world usage (notably charging/engine-dominant operation), highlighting the need to complement type-approval metrics with usage-sensitive indicators when evaluating PHEV compliance in practice.

Keywords: PHEV; vehicles; OBFCM; real-world CO_2; battery capacity; non-linear modeling; WLTP

1. Introduction

Plug-in hybrid electric vehicles (PHEVs) are positioned as an important transitional technology in European decarbonization strategies because, in principle, a traction battery can displace engine operation over a substantial share of daily driving while preserving long-distance capability [1,2]. In regulatory accounting, PHEV CO_2 performance is largely inferred from type-approval procedures (e.g., WLTP) combined with assumptions about how frequently the vehicle operates electrically in real use [3,4]. This framing implies a simple intuition: increasing battery capacity should increase electric driving and reduce real-world CO_2 [5,6]. From a vehicle perspective, traction battery capacity is only one element of the PHEV system and is tightly coupled to market segment, vehicle mass and performance targets, powertrain calibration, and—most importantly—real-world charging and operating regimes. Therefore, a simple comparison of vehicles with different battery sizes can conflate hardware effects with systematic differences in who buys the vehicle and how it is used. In this study we use type-approval (declared) CO_2 as a standardized reference and focus on the test-to-reality gap as a compliance-relevant indicator of real-world PHEV performance in use.

Check for updates

Academic Editor: Weixiang Shen

Received: 9 February 2026
Revised: 5 March 2026
Accepted: 13 March 2026
Published: 15 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Large-scale monitoring evidence has increasingly challenged that intuition by documenting systematic gaps between type-approval expectations and real-world outcomes for contemporary European PHEV fleets [7,8]. In prior OBFCM-based analyses, the distribution of gap% was found to be heavily right-skewed with an average around 300%, indicating that real-world fuel consumption and CO_2 emissions often exceed test-cycle expectations by a wide margin [9,10]. Importantly for this study, battery capacity shows only a weak association with real-world CO_2 and exhibits a weak but positive correlation with gap%, i.e., larger batteries are not automatically linked to smaller gaps.

These observations raise an unresolved question with direct implications for compliance analytics and policy: does traction battery capacity contain an independent, causal signal for the test-to-reality CO_2 gap, or does it mainly act as a proxy for unobserved factors such as market segment, vehicle mass and performance targets, powertrain architecture, and real-world usage regimes. Because battery size is strongly entangled with segmentation [11,12] (e.g., premium or larger vehicles tending to carry larger batteries), simple bivariate relationships cannot separate "battery effect" from systematic differences in who buys the vehicle and how it is operated. A credible assessment therefore requires modeling strategies that explicitly control for segmentation and manufacturer/model heterogeneity while accounting for usage-related mechanisms that are observable in fleet monitoring data [13,14]. In Europe, regulatory updates such as the Euro 6e amendment aim to address these discrepancies by adjusting the utility factor (UF) curve used in type-approval to better reflect real-world usage, acknowledging that PHEVs often operate with far lower electric driving shares than officially assumed [15]. Similarly, studies in the United States based on data from Fuelly and the California Bureau of Automotive Repair indicate that real-world electric drive shares are 26–56% lower than EPA label values, leading to fuel consumption 42–67% higher than certified figures [16]. These findings underscore a systemic challenge: PHEV performance is highly dependent on actual usage patterns—especially charging behavior—rather than nominal battery capacity alone. While prior research has focused on regulatory adjustments and aggregated fleet gaps, the role of battery capacity as an explanatory variable remains ambiguous.

This study addresses that gap using data from the Joint Research Centre's On-Board Fuel and energy Consumption Monitoring (OBFCM) database for light-duty M1 vehicles, concatenated for 2021–2023, where each record represents a unique vehicle and—when multiple OBFCM readouts exist for one vehicle—the most recent readout is retained. From this source we construct a "true PHEV" analytical sample using a minimum traction battery capacity threshold and quality filters, yielding 457,555 vehicles (and 452,872 observations in the fully specified regressions after excluding missing proxy variables) across 14 manufacturers. Methodologically, we quantify how the apparent battery–gap relationship changes along a nested fixed-effects ladder (segment, monitoring year, manufacturer), test whether non-linear battery terms add explanatory power (B-splines and partial-residual diagnostics), and estimate segment-dependent marginal battery slopes to directly evaluate heterogeneity and potential sign reversals. Finally, we perform robustness checks using model identifier fixed effects (MS_Cn) with standard errors clustered by MS_Cn to assess whether any residual battery signal remains once model-level heterogeneity is absorbed, thereby operationalizing and testing the "battery as proxy variable" interpretation in a large harmonized fleet dataset. This study asks whether traction battery capacity contains an independent explanatory signal for the test-to-reality CO_2 gap in European PHEVs, or whether the observed battery–gap association mainly reflects confounding by market segmentation and real-world usage regimes captured in OBFCM data. We hypothesize that once segment/year/manufacturer heterogeneity and OBFCM-derived usage proxies are controlled for, the apparent battery effect will attenuate substantially, consistent with

interpreting battery capacity primarily as a proxy variable rather than a universal lever of in-use CO_2 performance.

2. Materials and Methods

This study used vehicle-level data from the European Commission Joint Research Centre (JRC) On-Board Fuel and Energy Consumption Monitoring (OBFCM) dataset for light-duty M1 vehicles, concatenated for monitoring years 2021–2023 [17]. Using OBFCM on-board monitoring data enables a large-scale, harmonized, vehicle-level assessment of the test-to-reality CO_2 gap and supports controlling for segment, monitoring year and manufacturer heterogeneity. However, OBFCM is not a trip-level dataset and does not include detailed contextual variables (e.g., route characteristics, ambient temperature, charging access, or user intent), so the estimated battery–gap relationships should be interpreted primarily as associations at fleet scale rather than as causal effects of battery capacity. In addition, OBFCM coverage can be uneven across manufacturers and segments; therefore, we report the sample composition and use fixed effects and robustness checks to mitigate composition-driven confounding. On-Board Fuel and Energy Consumption Monitoring (OBFCM) is a standardized system installed in modern vehicles to record real-world fuel and energy usage. It collects detailed data on parameters such as distance traveled, fuel consumption, CO_2 emissions, and the operation of electric powertrains, providing a transparent view of actual vehicle performance beyond laboratory type-approval tests [18,19]. Each record represents a unique vehicle; when multiple OBFCM readouts were available for a given vehicle within the observation period, only the most recent readout was retained to avoid repeated counting and to better reflect the latest in-use state of the vehicle.

The methodological framework, illustrated in Figure 1, proceeds through the numbered steps (1–7). Step 1 uses the OBFCM database as the data source, providing vehicle-level real-world fuel/energy use, CO_2 emissions, and powertrain operation counters for a large PHEV fleet. Step 2 summarizes the construction of the core analytical variables. In Step 3, we define the dependent variable as the test-to-reality CO_2 gap% (RW_CO2 − TACO2)/TACO2, specify battery capacity (kWh) as the key explanatory variable, derive the engineered usage proxies (EUR, HI, EDE, ELP), and introduce categorical controls via fixed effects (segment, monitoring year, manufacturer; and MSCn in robustness checks) to account for structural and unobserved heterogeneity. In Step 4, we implement the statistical modeling strategy (nested OLS specifications, cubic B-splines for non-linearity checks, and interaction terms for heterogeneity). Steps 5–6 summarize how model outputs are translated into key findings (battery-effect attenuation after adding usage proxies, the dominant explanatory role of usage intensity, and segment-specific heterogeneity, including robustness to model-level fixed effects). Finally, Step 7 synthesizes these results into the main conclusion that battery capacity primarily acts as a proxy for segmentation and real-world usage rather than a direct lever of fleet-average CO_2 gap%. To keep the analysis closely connected to real vehicles, we interpret the engineered OBFCM indicators as usage descriptors of PHEV operation: EUR summarizes how much of charge-depleting driving is completed with the engine off, HI captures the intensity of engine-involved hybrid operation over lifetime distance, EDE reflects cumulative charging energy per kilometer, and ELP (RWFC/TAFC) approximates engine-dominant operation relative to the type-approval benchmark. These quantities are therefore not abstract mathematical features; they operationalize key mechanisms—charging frequency and engine dominance—that determine whether a given PHEV delivers low real-world CO_2 in practice.

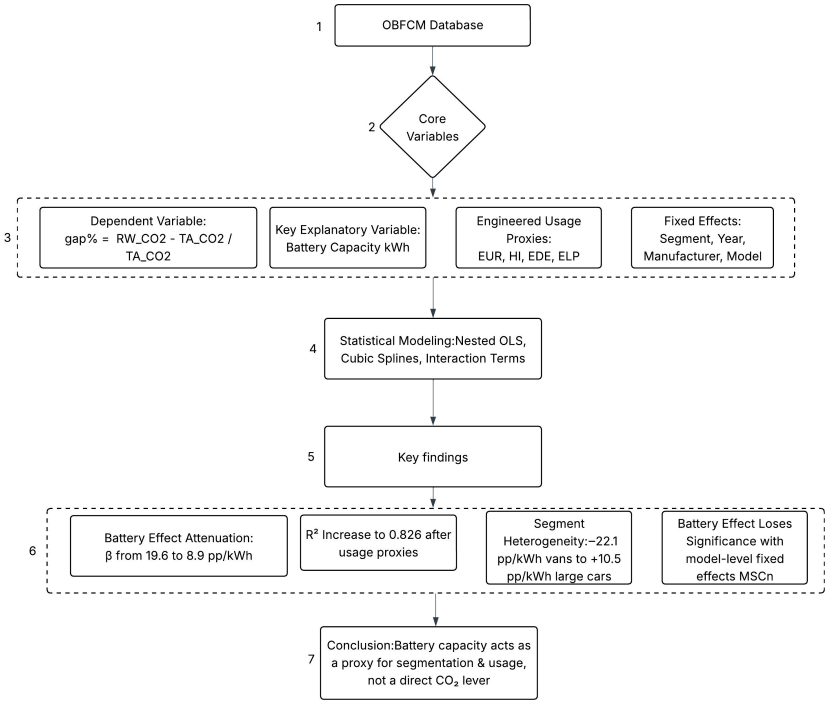


Figure 1. General workflow of the research.

A nested modeling strategy is then employed using Ordinary Least Squares (OLS) regression. Models sequentially add fixed effects and usage proxies to isolate the battery signal. Supplementary analyses include cubic B-splines to test for non-linear battery effects and segment-interaction terms to estimate heterogeneous marginal slopes.

The study demonstrates that traction battery capacity is primarily a proxy variable for market segmentation and, crucially, for real-world usage intensity (charging behavior and engine dominance). It does not serve as a universal, independent lever for real-world CO_2 performance. The findings underscore the importance of usage-based metrics—derivable from monitoring systems like OBFCM—over simple hardware specifications like battery size for effective compliance analytics and policy design targeting real-world PHEV decarbonization.

The raw database was assembled from multiple manufacturer-specific extracts and harmonized into a single analytical table including vehicle identifiers, monitoring year, brand, segment, type-approval reference values, and real-world OBFCM measurements. The core real-world variables used in this work include real-world fuel consumption (RW_FC), real-world CO_2 emissions (RW_CO2), and OBFCM distances describing charge-depleting operation with engine off and engine on, charge-increasing operation, total lifetime distance, and cumulative energy into the traction battery. Type-approval reference fuel consumption (TA_FC) was used to compute an engine-load proxy and to provide a consistent benchmark against real-world operation.

To focus the analysis on "true" plug-in hybrids and to ensure data validity, we applied a sequence of sample-selection criteria aligned with OBFCM-based fleet analyses. First, PHEVs were defined using a minimum traction battery capacity threshold of 1.56 kWh, which excludes conventional non-plug-in hybrids and mild hybrids while retaining vehicles with plug-in-capable battery systems. Second, quality filters were applied to remove implausible or extreme values likely reflecting data errors or non-representative operation: gap% was restricted to −100 to 1000, and RW_CO2 was restricted to 0 to 500 g/km. After applying these criteria, the final analytical cohort comprised 457,555 PHEVs, with

452,872 observations available for the fully specified regressions after listwise deletion of missing values in engineered proxies.

The primary dependent variable was the test-to-reality CO_2 gap expressed as a percentage (gap%), computed consistently with OBFCM literature as the relative difference between real-world and type-approval CO_2 (or corresponding fuel-consumption-based reference), i.e., a positive value indicates that real-world emissions exceed the type-approval expectation. Battery capacity was treated as the main explanatory variable of interest, entered as a continuous predictor and mean-centered (batt_c) for interpretability and numerical stability. In addition, we engineered usage-related proxies from OBFCM fields to represent key mechanisms linking driver behavior and operating conditions to real-world emissions: EUR (electric-mode utilization ratio within charge-depleting operation), HI (hybridization intensity capturing charge-depleting engine-on plus charge-increasing distance relative to lifetime distance), EDE (energy into battery per kilometer), and ELP (a proxy for engine-dominant operation defined as the ratio RW_FC/TA_FC, capped to reduce the influence of extreme outliers).

The statistical analysis was based on ordinary least squares (OLS) regression. To test whether battery capacity provides an independent signal or mainly acts as a proxy for segmentation and usage, we estimated a nested sequence of models that progressively add controls and fixed effects: (i) a battery-only baseline, (ii) segment and monitoring year fixed effects, (iii) manufacturer fixed effects, and (iv) engineered usage proxies. Heteroskedasticity-robust standard errors (HC3) were used as the default inference approach for baseline specifications. The statistical analysis was based on ordinary least squares (OLS) regression. To test whether battery capacity provides an independent signal or mainly acts as a proxy for segmentation and usage, we estimated a nested sequence of models that progressively add controls and fixed effects: (i) a battery-only baseline, (ii) segment and monitoring year fixed effects, (iii) manufacturer fixed effects, and (iv) engineered usage proxies. Heteroskedasticity-robust standard errors (HC3) were used as the default inference approach for baseline specifications. Because the dependent variable (gap%) is strongly right-skewed and includes negative values after quality screening (-96.6 to 999.96), we complemented OLS (HC3) with two robustness checks. First, we estimated an alternative specification using a shifted log transformation of the dependent variable, $\log(\text{gap} - \min(\text{gap}) + 1)$, which is well-defined for all observations and reduces the leverage of extreme right-tail outcomes. Second, we estimated quantile regressions ($\tau = 0.25, 0.50, 0.75$) for the fully controlled specification to test whether the conditional battery–gap association is concentrated in the upper tail or persists across the distribution. To evaluate whether the conditional battery–gap relationship departs from linearity, we additionally estimated models using cubic B-splines for battery capacity (df = 3–5) and used partial-residual diagnostics aggregated by battery-capacity deciles. Finally, to assess robustness to unobserved model-level heterogeneity, we estimated an alternative specification including model identifier fixed effects (MS_Cn) and computed cluster-robust standard errors clustered by MS_Cn, which is appropriate when residual correlation is expected within model identifiers. All computations were performed in Python using standard data-science libraries (pandas/numpy) and econometric routines from statsmodels.

Formally, the regulatory gap was computed as:

$$gap\% = \frac{RW_CO2 - TA_CO2}{TA_CO2} \times 100 \quad (1)$$

where

RW_CO2—real-world CO_2,

TA_CO2—corresponding type-approval (WLTP) CO_2 reference value.

Positive values therefore indicate underestimation of real-world emissions by the type-approval benchmark. To reduce the influence of implausible records and ensure comparability, gap% was restricted to the interval −100 to 1000 and RW_CO2 to 0–500 g/km, consistent with OBFCM-based quality screening practices.

Engineered proxies were computed directly from OBFCM distance and energy counters to represent use intensity and powertrain operating mode. Electric-mode utilization ratio (EUR, %) was defined within charge-depleting operation:

$$EUR = \frac{d_{CD,eng-off}}{d_{CD,eng-off} + d_{CD,eng-on}} \times 100 \quad (2)$$

where

$d_{CD,eng-off}$; $d_{CD,eng-on}$—correspond to OBFCM charge-depleting distances with engine off and engine on.

Hybridization intensity (HI, %) was defined as:

$$HI = \frac{d_{CD,eng-off} + d_{CI}}{d_{life}} \times 100 \quad (3)$$

where

d_{CI}—charge-increasing operation distance,

d_{life}—total lifetime distance.

Energy-to-distance efficiency (EDE, kWh/km) was computed as:

$$EDE = \frac{E_{into\ batt}}{d_{life}} \quad (4)$$

where

$E_{intobatt}$—cumulative energy into the battery (kWh).

The engine-load proxy (ELP, unitless) was computed from the ratio of real-world to type-approval fuel consumption and capped to limit the leverage of extreme observations, i.e., ELP = min(RW_FC/TA_FC,5), with larger values indicating more engine-dominant operation.

For all engineered variables, denominators were required to be strictly positive; observations with zero/negative denominators were set to missing for that derived metric. Missing values in raw OBFCM counters were treated conservatively (e.g., distance/energy counters set to zero only when consistent with the counter semantics) and the fully specified regression sample was obtained via listwise deletion across gap%, battery capacity, and engineered proxies. Battery capacity was mean-centered (batt_c) prior to model fitting to improve interpretability and numerical stability in specifications that included non-linear terms (splines).

3. Results

The final analytical sample comprised 457,555 unique PHEVs after applying the battery capacity threshold (≥1.56 kWh) and quality filters (gap% restricted to [−100, 1000] and $RWCO_2$ to [0, 500] g/km). Table 1 presents the descriptive statistics for key variables. The sample was dominated by upper medium cars (41.7%) and lower medium cars (37.0%), with large cars accounting for 19.9% and medium vans for 1.3% of observations. The majority of vehicles were monitored in 2021 (55.8%), with 2022 and 2023 contributing 35.1% and 9.2%, respectively. Battery capacity in the PHEV sample ranged from 1.56 to 21.6 kWh with a mean of 13.0 kWh (SD = 2.38 kWh). The distribution was strongly left-skewed, with 87.8% of the full raw database showing zero battery capacity values (conventional hybrids excluded from analysis), and the PHEV-only distribution concentrated between 11.6 and

14.1 kWh (IQR). The test-to-reality CO_2 gap exhibited substantial variation (mean = 300.1%, SD = 170.6%), with the median gap at 273.8%, indicating that real-world emissions typically exceeded type-approval expectations by nearly threefold.

Table 1. Descriptive statistics for PHEV analytical sample (n = 457,555).

Variable	Mean	SD	Min	Q25	Median	Q75
Dependent Variable						
gap% (test-to-reality CO_2 gap)	300.1	170.6	−96.6	179.8	273.8	391.5
Key Explanatory Variable						
Battery capacity (kWh)	13.04	2.38	1.56	12	12.86	13.61
Usage Proxies						
EUR (electric utilization ratio, %)	69.8	19.7	0	55.6	72.2	86.4
HI (hybridization intensity, %)	19.4	16.7	0	6.2	15.1	28.3
EDE (energy-to-distance, kWh/km)	0.07	0.056	0	0.022	0.063	0.107
ELP (engine-load proxy, unitless)	3.67	1.13	0.03	2.8	3.74	4.91
Control Variables						
$RWCO_2$ (g/km)	108.3	58.4	0	64.3	99.7	143

The engineered usage proxies showed (Table 1) considerable variability across the fleet. Electricmode utilization ratio (EUR) averaged 69.8% within charge-depleting operation (SD = 19.7%), with an interquartile range from 55.6% to 86.4%. Hybridization intensity (HI) averaged 19.4% (SD = 16.7%), reflecting the proportion of lifetime distance driven in engine on charge-depleting or charge-increasing modes. Energy-to-distance efficiency (EDE) averaged 0.070 kWh/km (SD = 0.056), while the engine-load proxy (ELP, defined as RWFC/TAFC and capped at 5.0) averaged 3.67 (SD = 1.13), indicating that real-world fuel consumption typically exceeded type-approval expectations by a factor of 3.7 for the engine-dominant operating component.

The manufacturer distribution was concentrated (Table 2) among premium European brands, with Volvo (33.9%) and BMW AG (25.3%) representing nearly 60% of the PHEV sample. This composition reflects both the European PHEV market structure during 2021–2023 and the availability of OBFCM data by manufacturer. The concentration in specific segments and brands underscores the importance of controlling for segmentation effects when assessing battery capacity associations with real-world gap%. Table 2 shows that the OBFCM PHEV sample is not an abstract population but a concrete fleet dominated by specific vehicle classes and brands. The prevalence of upper/lower-medium cars and the strong concentration in a few manufacturers imply that the observed battery–gap relationship can be driven by systematic differences in vehicle design choices (mass, performance targets, powertrain strategies) and customer use patterns across segments and brands. This fleet composition motivates the fixed-effects strategy used below to avoid attributing segment- or brand-specific usage regimes to battery capacity.

Table 2. Sample composition by segment, monitoring year, and manufacturer (n = 457,555).

Category	*n*	%
Market Segment		
Upper Medium Car	190,910	41.7
Lower Medium Car	169,482	37
Large Car	91,093	19.9
Medium Van	6070	1.3

Table 2. *Cont.*

Category	*n*	%
Monitoring Year		
2021	255,078	55.8
2022	160,434	35.1
2023	42,043	9.2
Manufacturer (Top 10)		
Volvo	155,351	33.9
BMW AG	115,729	25.3
SEAT	40,070	8.8
Volkswagen	36,022	7.9
Audi AG	31,860	7
Škoda	19,430	4.2
Mazda	17,923	3.9
Toyota	15,603	3.4
Fiat Group	12,677	2.8
Opel Automobile	7621	1.7

3.1. Separating Vehicle Class and Usage from Battery Size

To evaluate whether battery capacity provides an independent signal for the test-to-reality CO_2 gap or primarily acts as a proxy for segmentation and usage patterns, we estimated a nested sequence of OLS regression models with progressively added controls. The nested specifications in Table 3 are designed to progressively compare "more similar vehicles": starting from a naïve across-fleet relationship (battery only), then accounting for vehicle class (segment), time (monitoring year), and manufacturer differences, and finally controlling for OBFCM-derived usage descriptors. This stepwise approach helps separate battery capacity as a hardware attribute from the confounding influence of vehicle segmentation and real-world operating regimes. Table 3 presents the model comparison, showing the evolution of the battery capacity coefficient and model explanatory power (R^2) across four specifications.

Table 3. Nested fixed-effects models: battery capacity coefficient attenuation and R^2 progression.

Model	*n*	R^2	Battery (pp/kWh)	*p*-Value
M0: Battery only	457,555	0.075	19.63	<0.001
M1: + Segment + Year FE	457,555	0.185	18.46	<0.001
M2: + Manufacturer FE	457,555	0.203	17.45	<0.001
M3: + Usage proxies (EUR, HI, EDE, ELP)	452,872	0.826	8.88	<0.001

Robustness checks addressing the strong right-skewness of gap% indicate that the main conclusions are not driven solely by a small number of extreme observations. In the shifted-log specification $\log(\text{gap} - \min(\text{gap}) + 1)$, the battery coefficient remains positive (0.0107; SE = 0.00011, n = 452,872). Quantile regressions also yield positive battery effects across the distribution: 23.76 (SE = 0.059) at $\tau = 0.25$, 25.93 (SE = 0.077) at $\tau = 0.50$, and 27.79 (SE = 0.116) at $\tau = 0.75$. Model M0, which regressed gap% on battery capacity alone, yielded a positive coefficient of 19.6 percentage points per kWh ($p < 0.001$), with an R^2 of only 0.075. This baseline model confirms the counterintuitive positive bivariate association: larger batteries are associated with higher test-to-reality gaps, contradicting the regulatory expectation that larger batteries should enable more electric driving and reduce real-world CO_2 emissions. Adding segment and monitoring year fixed effects (M1) increased R^2 to 0.185, indicating that market segmentation explains an additional 11 percentage points of gap% variation. The battery coefficient attenuated slightly to 18.5 pp/kWh, suggesting

that part of the battery-gap association is confounded by segment-level differences in vehicle characteristics and usage patterns. Further addition of manufacturer fixed effects (M2) raised R^2 to 0.203 and reduced the battery coefficient to 17.5 pp/kWh, reflecting manufacturer-specific heterogeneity in powertrain design, calibration strategies, and customer profiles. The most substantial change occurred in Model M3, which incorporated the four engineered usage proxies (EUR, HI, EDE, ELP) alongside all fixed effects. R^2 increased dramatically to 0.826, indicating that real-world usage intensity explains the vast majority of gap% variation once segmentation and manufacturer effects are controlled. Critically, the battery capacity coefficient attenuated by 54.7% from the baseline, declining to 8.9 pp/kWh ($p < 0.001$). This attenuation demonstrates that a large portion of the apparent battery-gap relationship is mediated by usage patterns: vehicles with larger batteries tend to be driven in more engine-dominant modes, offsetting any potential electric-range advantage. Among the usage proxies in Model M3, the engine-load proxy (ELP) exhibited the strongest association with gap% ($\beta = 121.1$, $p < 0.001$), confirming that the ratio of real-world to type approval fuel consumption is the dominant predictor of the test-to-reality gap. Electric mode utilization ratio (EUR) showed a negative association ($\beta = 0.22$, $p < 0.001$), indicating that higher electric-mode usage within charge-depleting operation is associated with marginally higher gaps when all other factors are held constant—a result likely reflecting that EUR itself is conditioned on charge-depleting events rather than overall trip mix. Energy-to-distance efficiency (EDE) exhibited a large negative coefficient ($\beta = -305.2$, $p < 0.001$), consistent with the interpretation that cumulative battery charging per kilometer driven reflects more frequent charging and electric operation, which partially offsets engine-dominant usage.

3.2. Non-Linear Battery Capacity Effects: Cubic Spline Analysis

To assess whether the conditional battery–gap relationship departs from linearity, we first examined partial residuals of gap% aggregated by battery-capacity deciles from the full usage-controlled model (M3). Battery capacity was divided into ten equal-count deciles between 1.56 and 21.6 kWh, and for each decile we computed the mean and interquartile range of the partial residuals, holding segment, monitoring year, manufacturer and usage proxies (EUR, HI, EDE, ELP) constant.

Figure 2 shows that the conditional relationship between battery capacity and the test-to-reality gap is clearly non-monotonic. The mean partial residual is slightly negative in the lowest two deciles (up to about 11.6 kWh, roughly −15 to −20 percentage points), becomes positive in the 11.6–13.0 kWh range (peaking around +15–20 pp), returns close to zero around 12.9–13.0 kWh, turns mildly negative again in the 13.0–13.8 kWh region (down to about −15 pp), and then rises sharply for the highest-capacity decile (14.2–21.6 kWh, about +35 pp). The shaded interquartile bands indicate that despite this structure, substantial within-decile variation remains, with IQR widths between roughly 40 and 70 percentage points.

Despite the visual evidence of non-monotonicity in the partial residuals, the incremental explanatory power of non-linear battery terms was modest. Allowing cubic B-splines with df = 3 increased R^2 from 0.826 to 0.828 (+0.2 percentage points), with df = 4 yielding $R^2 = 0.829$ (+0.3 pp) and df = 5 yielding $R^2 = 0.829$ (no further improvement). These results indicate that while the conditional battery-gap relationship is not strictly linear, the non-linear component explains less than 0.5% of gap% variation once usage proxies and fixed effects are included.

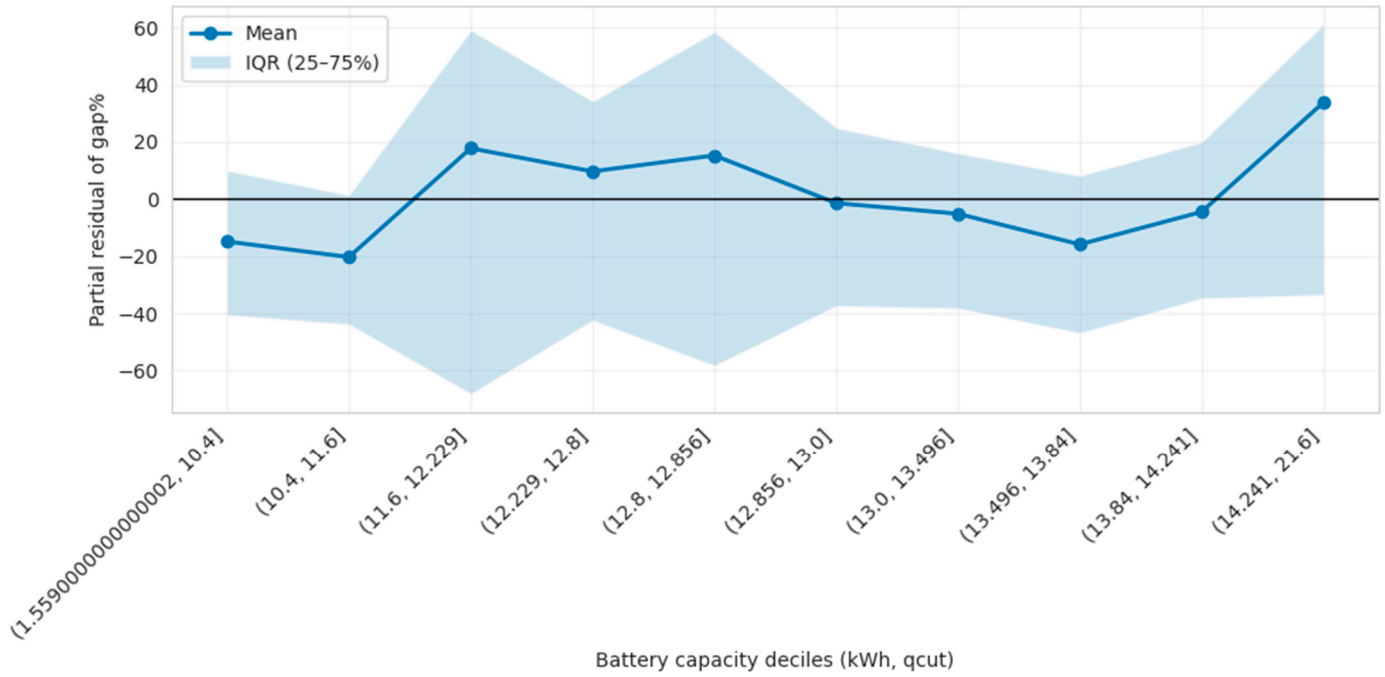


Figure 2. Partial residuals (battery deciles).

Figure 3 illustrates the predicted gap% as a function of battery capacity using the cubic spline model (df = 4), with all usage proxies and categorical controls held at their median or modal values (EUR = 72.2%, HI = 15.1%, EDE = 0.063 kWh/km, ELP = 3.74, segment = Upper Medium Car, year = 2021, manufacturer = Volvo). The prediction curve confirms the nonmonotonic shape observed in the partial residuals: predicted gap% peaks around 13–14 kWh at approximately 315%, declines to a local minimum around 17–18 kWh (approximately 280%), and then increases sharply for the largest batteries (>20 kWh, exceeding 410%). This U-shaped pattern suggests that the marginal effect of battery capacity on gap% reverses sign depending on the battery size range, likely reflecting the interaction between battery capacity, vehicle segment, and real-world usage regimes.

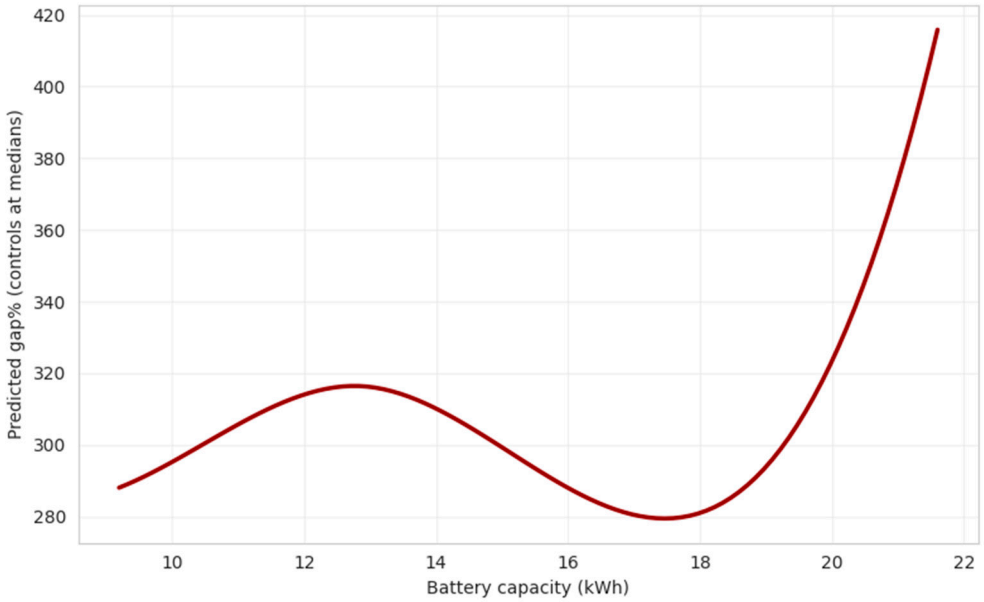


Figure 3. Spline prediction curve (controls at medians).

3.3. Usage Proxy Relationships: Electric Utilization and Gap%

The engineered usage proxies provide direct insight into how real-world operating behavior mediates the test-to-reality gap. Figure 4 shows a density visualization (hexbin) of EUR (electric-mode utilization ratio within charge-depleting operation) versus gap%, with segment-specific binned mean curves overlaid to improve interpretability in this large OBFCM sample. The overall negative association is evident: higher EUR values are associated with lower average gaps, while low EUR values are linked to substantially higher gaps and greater dispersion. The vertical accumulation near EUR $\approx$ 100% reflects vehicles for which the OBFCM charge-depleting engine-on counter is close to zero, i.e., charge-depleting operation is recorded almost exclusively with the engine off. At the same time, the remaining dispersion at similar EUR values is expected because EUR is conditioned on charge-depleting operation and does not capture how frequently charge-depleting operation occurs in overall use (e.g., trip-length mix, charging frequency, and the share of charge-sustaining driving). Table 4 therefore complements Figure 4 by quantifying the monotonic decline in average gap% across EUR deciles despite point-level variability.

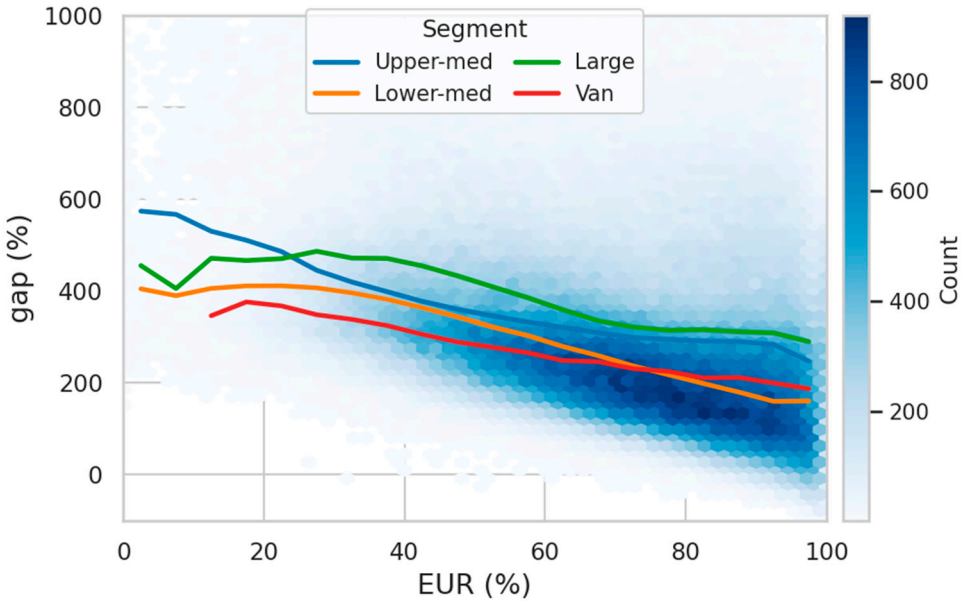


Figure 4. Relationship between electric utilization ratio (EUR) and the test-to-reality CO_2 gap (gap%). Hexbin shading indicates observation density; solid lines show segment-wise binned mean gap% across EUR.

Table 4. Mean and median gap% by electric-mode utilization ratio (EUR) deciles.

EUR Decile (Range, %)	*n*	Mean Gap%	Median Gap%
D1: 0.0–41.7	45,288	420.7	391.1
D2: 41.7–51.6	45,287	372.3	342.6
D3: 51.6–59.3	45,287	340.1	309.8
D4: 59.3–66.1	45,287	312.6	280.7
D5: 66.1–72.2	45,287	290.7	256.7
D6: 72.2–78.1	45,287	274.5	235.9
D7: 78.1–83.7	45,287	264.6	222.7
D8: 83.7–89.0	45,287	256	213.3
D9: 89.0–94.1	45,287	245.5	204.5
D10: 94.1–100.0	45,288	223.1	190.6

The Pearson correlation between EUR and gap% was -0.34 ($p < 0.001$), confirming a moderate negative association. To further quantify this relationship, Table 4 presents

mean and median gap% by EUR deciles. Vehicles in the lowest EUR decile (EUR $\leq$ 41.7%) exhibited a mean gap% of 420.7% (median = 391.1%), while those in the highest decile (EUR $\geq$ 94.1%) had a mean gap% of 223.1% (median = 190.6%). This 197-percentage-point difference in mean gap% across the EUR distribution underscores the dominant role of real-world charging and electric-mode usage in determining PHEV compliance outcomes.

The monotonic decline in gap% across EUR deciles demonstrates that the frequency and intensity of electric-mode operation—not battery capacity per se—is the primary 3.4 Usage Proxy Relationships: Electric Utilization and Gap% determinant of real-world PHEV CO_2 performance. Vehicles with large batteries but low charging frequency (low EUR) perform similarly to or worse than vehicles with smaller batteries but high charging frequency (high EUR), reinforcing the interpretation that battery capacity acts as a proxy for usage patterns rather than a direct causal lever.

3.4. Segment-Specific Marginal Battery Slopes: Heterogeneity and Sign Reversals

Given the strong entanglement between battery capacity and market segmentation, we estimated segment-specific marginal battery slopes by including battery $\times$ segment interaction terms in the full model (M3 specification). Table 5 and Figure 5 present the estimated marginal effects of battery capacity on gap% for each segment, along with 95% confidence intervals.

Table 5. Segment-specific marginal battery slopes on gap% (pp per kWh).

Segment	Slope (pp/kWh)	SE	95% CI Lower	95% CI Upper
Medium Van	−22.15	0.30	−22.73	−21.56
Lower Medium Car	0.97	0.11	0.77	1.18
Upper Medium Car	7.13	0.10	6.93	7.34
Large Car	10.49	0.09	10.31	10.67

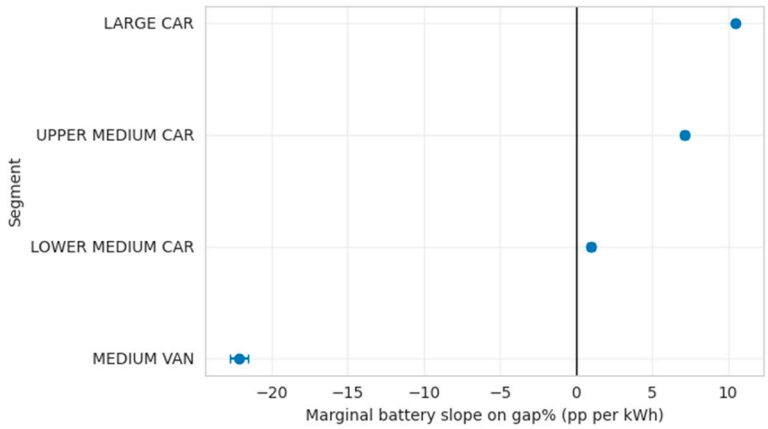


Figure 5. Segment-specific marginal battery slopes (forest point plot).

Based on Figure 5 the marginal battery slope for medium vans was −22.1 pp/kWh (95% CI: −22.7, −21.6), indicating that within this segment, larger batteries are associated with substantially lower test-to-reality gaps. This negative relationship is consistent with fleet or commercial usage patterns where vans are charged regularly and operated on predictable routes, allowing larger batteries to deliver their intended electric-range benefit. In contrast, the marginal slope for large cars was +10.5 pp/kWh (95% CI: 10.3, 10.7), indicating that larger batteries in the premium large-car segment are associated with higher gaps. This counterintuitive result likely reflects a combination of factors: premium large cars with bigger batteries tend to be heavier, more powerful, and driven by users with 3.5 Segment-Specific Marginal Battery Slopes: Heterogeneity and Sign Reversals longer trip

distances and lower charging frequency, leading to predominantly engine-based operation despite the larger battery capacity. Upper medium cars exhibited an intermediate positive slope of +7.1 pp/kWh (95% CI: 6.9, 7.3), while lower medium cars showed a near-zero marginal effect (+1.0 pp/kWh, 95% CI: 0.8, 1.2). The progression from strongly negative (medium vans) through near-zero (lower medium cars) to strongly positive (large cars) demonstrates that the battery-gap relationship is not universal but contingent on segment-specific usage regimes, vehicle characteristics, and customer profiles. The sign reversal across segments is a critical finding for compliance analytics and policy design. It indicates that battery capacity alone cannot serve as a reliable proxy for real world PHEV CO_2 performance without accounting for segmentation and usage context. In particular, policies or incentive structures that reward larger batteries without conditioning on real-world charging behavior may inadvertently favor segments and usage patterns where larger batteries do not translate into lower emissions. This heterogeneity is consistent with segment-specific vehicle use cases. In medium vans, PHEVs are more likely to operate on regular duty cycles with predictable charging opportunities, so larger batteries can translate into more electric operation and lower gaps. In contrast, in large cars, higher mass/performance and longer-trip usage can lead to more frequent engine operation, so larger batteries do not automatically reduce real-world CO_2 relative to type-approval.

3.5. Robustness Check: Model-Level Fixed Effects and Clustered Standard Errors

To further test the battery-as-proxy interpretation, we estimated an alternative specification that absorbs all model-level heterogeneity by including fixed effects for each unique model identifier (MSCn, n = 209 models). Standard errors were clustered by MSCn to account for within-model correlation in residuals. This specification effectively isolates variation in gap% and battery capacity within models that share the same powertrain architecture, market positioning, and expected customer base, providing a stringent test of whether battery capacity retains explanatory power once model-level confounding is removed.

The results of the MSCn fixed-effects model showed that the battery capacity coefficient remained positive (β = 8.88 pp/kWh in the manufacturer-FE model) but became statistically non-significant (p = 0.085) when MSCn fixed effects and clustered standard errors were applied. In contrast, the usage proxies retained their strong associations: EUR (β = 0.23, $p < 0.001$), HI (β = 0.10, $p < 0.001$), EDE ($\beta = -305.2$, $p < 0.001$), and especially ELP (β = 121.1, $p < 0.001$) remained highly significant and substantively unchanged in magnitude. This robustness check reinforces the interpretation that battery capacity's apparent association with gap% is largely spurious, driven by confounding with model-level characteristics (e.g., powertrain calibration, mass, aerodynamics, target customer segment) rather than a direct causal effect of battery size on real-world charging frequency or electric-mode operation. Once these model-level confounders are absorbed via MSCn fixed effects, the residual battery variation within models (e.g., minor capacity differences due to degradation or specification updates) has no detectable association with gap%. Meanwhile, usage intensity—as captured by the engineered proxies—remains the dominant and robust predictor of PHEV real-world CO_2 performance.

4. Discussion

The central finding of this study is that traction battery capacity provides only a weak independent signal for PHEV test-to-reality CO_2 gaps once segmentation, manufacturer heterogeneity, and real-world usage patterns are accounted for. In the baseline bivariate specification, larger batteries are associated with larger gaps (β = 19.6 pp/kWh, R^2 = 0.075). However, this association attenuates by ~55% to 8.9 pp/kWh after adding usage prox-

ies (R^2 = 0.826) and becomes statistically non-significant in the robustness specification with model identifier fixed effects (MSCn) and clustered inference (p = 0.085). Together, these results support interpreting battery capacity primarily as a proxy for correlated, largely unobserved model/segment positioning and in-use regimes rather than as a direct determinant of charging frequency or engine-off operation at fleet scale.

This proxy interpretation is consistent with how battery capacity is embedded in the OBFCM fleet composition and segment structure. Larger batteries are concentrated in upper-medium and large-car segments, where vehicle mass/performance targets and customer use (e.g., longer trip distances, lower charging regularity) can promote engine-dominant operation, while in medium vans larger batteries may translate more directly into electric use under predictable routes and disciplined charging. The segment interaction results reinforce this heterogeneity by showing sign reversals in the marginal battery slope across segments (negative in medium vans, near-zero in lower-medium cars, positive in large cars), implying there is no universal "bigger battery → smaller gap" relationship that holds across vehicle classes.

A second key finding is that usage-related indicators dominate the explanatory power for the gap once structural heterogeneity is controlled. In the nested models, adding the engineered usage proxies (EUR, HI, EDE, ELP) produces the largest improvement in fit (R^2 rising to 0.826) and strongly attenuates the battery coefficient, indicating that the observed battery–gap association is largely mediated by real-world operating behavior captured by OBFCM counters. Among these proxies, ELP (defined from the ratio RWFC/TAFC and capped to limit outliers) exhibits the strongest association with gap, highlighting that engine-dominant operation relative to the type-approval benchmark is the main empirical driver of compliance shortfalls in this dataset. This magnitude is consistent with recent OBFCM-based fleet evidence; for example, Ariadne Projekt estimated 6.0–6.2 L/100 km real-world fuel consumption versus 1.0–1.25 L/100 km WLTP, and JRC/EEA reported approximately 3.5× multipliers for 2021 PHEVs [20]. The descriptive EUR patterns (including its negative association with gap and the monotonic decline in average gap across EUR deciles) are directionally consistent with the interpretation that electric utilization within charge-depleting operation is linked to lower gaps, while also reflecting that EUR is a conditional metric based on recorded charge-depleting activity rather than a complete descriptor of trip mix.

The non-linearity checks further clarify what battery capacity adds once usage and fixed effects are included. Partial residual diagnostics by battery capacity deciles suggest a non-monotonic conditional pattern, yet the incremental gain from allowing cubic B-splines is small (R^2 increases by about 0.002–0.003 relative to the linear specification), implying that non-linear battery terms contribute little additional explanatory power beyond the controls. This combination of visible structure but minimal fit improvement is consistent with the idea that observed "non-linearity" largely reflects discrete battery size clusters tied to segment/model transitions rather than smooth technological gradients within comparable vehicles. Substantial within-bin dispersion in the partial residuals (wide interquartile ranges) also indicates meaningful residual heterogeneity at the vehicle level that is not captured by cumulative OBFCM summaries alone.

These results align with prior OBFCM-based evidence that real-world PHEV outcomes are driven more by usage (charging frequency, trip patterns, and engine-dominant operation) than by nominal battery size, and they extend that evidence by quantifying how the battery coefficient changes across a nested fixed-effects ladder and by documenting segment-specific sign reversals. The mean fleet-level gap observed here also echoes concerns raised by European NGOs and agencies that PHEVs may deliver only a fraction of the expected on-road CO_2 benefits under current type-approval assumptions. This

strengthens the case for systematic use of OBFCM-based monitoring to increase transparency and accountability in real-world compliance assessments [21]. In line with recent policy-oriented analyses, any future role for PHEVs (e.g., a limited post-2035 allowance) would require conditioning eligibility on verified real-world performance criteria measured via OBFCM [22].

Our findings are directly relevant to ongoing revisions of PHEV regulatory assumptions, including the utility-factor (UF) methodology and the implementation of Euro 6e/7 and future EU fleet CO_2 standards. Because traction battery capacity largely reflects market segmentation and real-world usage regimes rather than an independent lever of in-use CO_2 performance, continued reliance on hardware-based parameters alone risks perpetuating systematic type-approval-to-real-world compliance gaps. Regulatory decision-making should therefore increasingly anchor PHEV assessment in OBFCM-derived, usage-sensitive metrics—both to inform UF calculations and to support segment-differentiated benchmarks that reflect observed charging behavior. Conditioning regulatory crediting on verified real-world operation can improve alignment between type-approval assumptions and on-road outcomes and support more credible enforcement of future fleet CO_2 limits.

Importantly, the analysis is cross-sectional at the vehicle level (each record represents a single OBFCM readout within the monitoring window), therefore the estimates should be interpreted as associations at fleet scale rather than causal effects of battery capacity. Because OBFCM provides cumulative on-board counters rather than trip-level records and lacks contextual covariates (e.g., charging access, route topology, ambient conditions), omitted variables may contribute to the dispersion observed at similar EUR values. In addition, OBFCM-based datasets can be affected by OBFCM-related biases, including uneven coverage across manufacturers/segments and monitoring years, differences in data completeness across brands, and selection effects introduced by quality filters and missing denominators in engineered proxies; these factors may influence both the magnitude and variability of the estimated gap. This interpretation is also consistent with regulatory work emphasizing utility-factor (UF) corrections to better align type-approval assumptions with observed real-world operation [23–25]. Future work could strengthen causal interpretation by linking OBFCM to trip-resolved telematics/GPS or by constructing panel structures from repeated readouts to track changes over time (e.g., degradation and evolving charging behavior) [26]. The findings underscore the need to integrate OBFCM evidence with complementary real-world and telematics data to refine policy-relevant modeling frameworks and strengthen the empirical basis for future PHEV regulatory assessments [27,28].

5. Conclusions

This study aimed to test whether traction battery capacity in PHEVs contains an independent signal for the test-to-reality CO_2 compliance gap, or whether it mainly acts as a proxy for vehicle segmentation and real-world operating regimes. Using European OBFCM vehicle-level monitoring data (2021–2023), we computed the test-to-reality gap as the percentage difference between real-world and type-approval CO_2 and constructed OBFCM-based usage proxies capturing electric utilization and engine-dominant operation (EUR, HI, EDE, and ELP). We then applied a nested econometric strategy based on OLS, sequentially adding segment, monitoring year, manufacturer fixed effects, and finally the engineered usage proxies; we additionally tested non-linearity using cubic splines, estimated segment-specific marginal battery slopes via interaction terms, and performed a robustness check with model identifier fixed effects (MSCn) and clustered inference.

Quantitatively, battery capacity shows a positive bivariate association with the gap (β = 19.6 pp/kWh; R^2 = 0.075), but this relationship attenuates by ~55% after introducing fixed effects and OBFCM usage proxies (β = 8.9 pp/kWh; R^2 = 0.826) and becomes

statistically non-significant once MSCn fixed effects with clustered inference are applied (p = 0.085). Segment-specific estimates reveal strong heterogeneity and sign reversals, ranging from −22.1 pp/kWh in medium vans to +10.5 pp/kWh in large cars, confirming that there is no universal battery–gap relationship across vehicle classes. Usage intensity dominates explanatory power: ELP (RWFC/TAFC, capped) is the strongest proxy for engine-dominant operation, and average gap declines markedly across EUR deciles (from 421% at EUR < 42% to 223% at EUR > 94%). Non-linear battery terms add only marginal explanatory power beyond the full control set, suggesting that battery size mainly reflects discrete segment/model differences rather than a smooth lever of real-world compliance.

From a regulatory and compliance analytics perspective, these results imply that battery capacity alone is an unreliable indicator of real-world PHEV CO_2 performance because it conflates hardware with segmentation and user behavior. OBFCM-derived, usage-sensitive indicators (e.g., proxies reflecting electric utilization and engine-dominant operation) provide substantially more policy-relevant information for identifying compliance shortfalls and for designing incentives or verification approaches that are robust across segments and manufacturers. Finally, because OBFCM provides cumulative counters rather than trip-level context, the findings should be interpreted as fleet-scale associations; future work linking OBFCM with trip-resolved telematics/GPS or panel readouts would help isolate causal mechanisms and temporal effects (e.g., degradation and evolving charging behavior).

Author Contributions: Conceptualization, P.K. and M.M.; methodology, M.M.; software, M.M.; validation, T.C., P.K. and M.M.; formal analysis, T.C.; investigation, M.M.; resources, M.M.; data curation, M.M.; writing—original draft preparation, M.M.; writing—review and editing, M.M.; visualization, M.M.; supervision, T.C.; project administration, P.K.; funding acquisition, T.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: This study uses the European Commission Joint Research Centre (JRC) OBFCM real-world dataset for M1 vehicles (monitoring years 2021–2023). The dataset is available via the JRC source described in Ref.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BEV	Battery Electric Vehicle
CD	Charge-Depleting (mode)
CI	Charge-Increasing (operation)
CO_2	Carbon Dioxide
CS	Charge-Sustaining (mode)
EDE	Energy-to-Distance Efficiency
ELP	Engine-Load Proxy
EUR	Electric Utilization Ratio
FE	Fixed Effects
GHG	Greenhouse Gas
HI	Hybridization Intensity
OBFCM	On-Board Fuel and energy Consumption Monitoring
OLS	Ordinary Least Squares
PEMS	Portable Emissions Measurement System
PHEV	Plug-in Hybrid Electric Vehicle
pp/kWh	Percentage points per kilowatt-hour

RDE	Real Driving Emissions
RW_CO_2	Real-World CO_2 emissions
RW_FC	Real-World Fuel Consumption
TA_CO_2	Type-Approval CO_2 emissions
TA_FC	Type-Approval Fuel Consumption
UF	Utility Factor
WLTP	Worldwide Harmonized Light Vehicles Test Procedure

References

1. Neves, S.A.; Marques, A.C. What has driven the adoption of BEV and PHEV in the EU? *Res. Transp. Bus. Manag.* **2025**, *60*, 101331. [CrossRef]
2. Pollák, F.; Vodák, J.; Soviar, J.; Markovič, P.; Lentini, G.; Mazzeschi, V.; Luè, A. Promotion of electric mobility in the European Union—Overview of project PROMETEUS from the perspective of co-hesion through synergistic cooperation on the example of the catching-up region. *Sustainability* **2021**, *13*, 1545. [CrossRef]
3. Liu, X.; Zhao, F.; Hao, H.; Chen, K.; Liu, Z.; Babiker, H.; Amer, A.A. From NEDC to WLTP: Effect on the Energy Consumption, NEV Credits, and Subsidies Policies of PHEV in the Chinese Market. *Sustainability* **2020**, *12*, 5747. [CrossRef]
4. Pavlovic, J.; Tansini, A.; Fontaras, G.; Ciuffo, B.; Otura, M.G.; Trentadue, G.; Bertoa, R.S.; Millo, F. *The Impact of WLTP on the Official Fuel Consumption and Electric Range of Plug-In Hybrid Electric Vehicles in Europe (No. 2017-24-0133)*; SAE Technical Paper; SAE International: Warrendale, PA, USA, 2017.
5. Pielecha, J.; Gis, W. Testing Exhaust Emissions of Plug-In Hybrid Vehicles in Poland. *Energies* **2024**, *17*, 6288. [CrossRef]
6. Pielecha, J.; Skobiej, K.; Kubiak, P.; Wozniak, M.; Siczek, K. Exhaust emissions from plug-in and HEV vehicles in type-approval tests and real driving cycles. *Energies* **2022**, *15*, 2423. [CrossRef]
7. Tansini, A.; Pavlovic, J.; Fontaras, G. Quantifying the real-world CO_2 emissions and energy consumption of modern plug-in hybrid vehicles. *J. Clean. Prod.* **2022**, *362*, 132191. [CrossRef]
8. Montazeri-Gh, M.; Pourbafarani, Z.; Mahmoodi-k, M. Comparative study of different types of PHEV optimal control strategies in real-world conditions. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2018**, *232*, 1597–1610. [CrossRef]
9. Mądziel, M.; Campisi, T. Real-World CO_2 Emissions of Plug-In Hybrid Vehicles: European Assessment Using On-Board Fuel Consumption Monitoring Data. *Energies* **2026**, *19*, 1165. [CrossRef]
10. Tansini, A.; Marin, A.L.; Suarez, J.; Aguirre, N.F.; Fontaras, G. Learning from the real-world: Insights on light-vehicle efficiency and CO_2 emissions from long-term on-board fuel and energy consumption data collection. *Energy Convers. Manag.* **2025**, *335*, 119816. [CrossRef]
11. Patil, T.-D.; Vinot, E.; Ehrenberger, S.; Trigui, R.; Redondo-Iglesias, E. Sensitivity analysis of battery aging for model-based PHEV use scenarios. *Energies* **2023**, *16*, 1749. [CrossRef]
12. Guozhi, P.; Tao, Z.; Guanhong, D.; Zekang, H.; Feifan, Y.; Shaobo, X. Battery sizing for plug-in hybrid electric vehicles considering social cost of carbon. *J. Energy Storage* **2025**, *131*, 117556. [CrossRef]
13. Zacharof, N.; Broekaert, S.; Grigoratos, T.; Bitsanis, E.; Fontaras, G. A real world assessment of European medium-duty vehicle emissions and fuel consumption. *Atmos. Environ. X* **2025**, *25*, 100307. [CrossRef]
14. Suarez, J.; Tansini, A.; Ktistakis, M.A.; Marin, A.L.; Komnos, D.; Pavlovic, J.; Fontaras, G. Towards zero CO_2 emissions: Insights from EU vehicle on-board data. *Sci. Total. Environ.* **2025**, *1001*, 180454. [CrossRef] [PubMed]
15. Dornoff, J. *Euro 6e: Changes to the European Union Light-Duty Vehicle Type-Approval Procedure; POLICY*; International Council on Clean Transportation: Washington, DC, USA, 2022.
16. Isenstadt, A.; Yang, Z.; Searle, S.; German, J. *Real World Usage of Plug-In Hybrid Vehicles in the United States*; International Council on Clean Transportation: Washington, DC, USA, 2022.
17. Suarez Corujo, J.; Tansini, A.; Ktistakis, M.; Laverde Marin, A.; Komnos, D.; Pavlovic, J.; Fontaras, G. *OBFCM Real-World Data from M1 Vehicles*; European Commission, Joint Research Centre (JRC): Brussels, Belgium, 2025.
18. Dimaratos, A.; Kontses, D.; Doulgeris, S.; Zacharof, N.; Keramidas, A.; Stamos, G.; Papageorgiou-Koutoulas, A.; Samaras, Z. *The Potential of On-Board Data Monitoring for the Characterization of Real-World Vehicle Fuel and Energy Consumption and Emissions (No. 2023-24-0113)*; SAE Technical Paper; SAE International: Warrendale, PA, USA, 2023.
19. Tapak, P.; Kocur, M.; Matej, J. On-Board Fuel Consumption Meter Field Testing Results. *Energies* **2023**, *16*, 6861. [CrossRef]
20. Plötz, P.; Gnann, T. *Real-World Fuel Consumption and Potential Future Regulation of Plug-In Hybrid Electric Vehicles in Europe–An Empirical Analysis of About One Million Vehicles*; Potsdam-Institut für Klimafolgen-forschung (PIK): Potsdam, Germany, 2025.
21. Plötz, P.; Link, S.; Ringelschwendner, H.; Keller, M.; Moll, C.; Bieker, G.; Dornoff, J.; Mock, P. *Real-World Usage of Plug-In Hybrid Vehicles in Europe: A 2022 Update on Fuel Consumption, Electric Driving, and CO_2 Emissions*; International Council on Clean Transportation: Washington, DC, USA, 2022.

22. Trasnport & Environemnt. Document: How Can fuel Consumption Meters Ensure Car CO_2 Limits Are Met On The Road. 2020. Available online: https://www.transportenvironment.org/articles/how-can-fuel-consumption-meters-ensure-car-co2-limits-are-met-road (accessed on 5 February 2026).
23. Zhao, Y.; Wang, F.; Zhang, S.; Wu, Y.; Hao, J. Decarbonization potential of Chinese PHEVs: Impact of increased battery range, power mix, and usage patterns. *Energy* **2025**, *331*, 136986. [CrossRef]
24. Mądziel, M. Phase-Specific Mixture of Experts Architecture for Real-Time NO_x Prediction in Diesel Vehicles: Advancing Euro 7 Compliance. *Energies* **2025**, *18*, 5853. [CrossRef]
25. Pielecha, J.; Woś, P.; Kuszewski, H.; Mądziel, M.; Krzemiński, A.; Kulasa, P.; Gis, W.; Piątkowski, P.; Sobczak, J. Real-Driving Emissions of Euro 2–Euro 6 Vehicles in Poland—17 Years of Experience. *Appl. Sci.* **2026**, *16*, 348. [CrossRef]
26. Mądziel, M.; Campisi, T. Predicting Auxiliary Energy Demand in Electric Vehicles Using Physics-Based and Machine Learning Models. *Energies* **2025**, *18*, 6092. [CrossRef]
27. Hamza, K.; Laberteaux, K.; Chu, K.-C. *An Approach for Estimating Contributions of Real-World Factors Towards Attained Well-to-Wheels Greenhouse Gas Emissions of Plug-In Hybrid Electric Vehicles (No. 2025-01-8543)*; SAE Technical Paper; SAE International: Warrendale, PA, USA, 2025.
28. Mancaruso, E.; Meccariello, G.; Rossetti, S. Investigation of Plug-In Hybrid Light Duty Commercial Vehicle in Real-World Conditions by Simulation. In *Conference on Sustainable Mobility*; SAE Technical Paper; SAE International: Warrendale, PA, USA, 2024.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.